



09-26-03

PATENT  
674558-2002

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

Applicants : SAITOU ET AL.  
Serial No. : 10/621,911  
Filed : JULY 17, 2003  
Examiner : TBA  
Group Art Unit : TBA  
For : GENES

745 Fifth Avenue  
New York, New York 10151

Mailing Label Number: EXPRESS MAIL  
EV3450130164S.

Date of Deposit: September 25, 2003  
I hereby certify that this paper or fee is being deposited with the  
United States Postal Service "Express Mail Post Office to  
Addressee" Service under 37 CFR 1.10 on the date indicated above  
and is addressed to: **Mail Stop 313(c), Commissioner for Patents,  
P.O. Box 1450, Alexandria, VA 22313-1450.**

Charles Jackson  
(Typed or printed name of person mailing paper or fee)

Charles Jackson  
(Signature of person mailing paper or fee)

Commissioner for Patents,  
P.O. Box 1450,  
Alexandria, VA 22313-1450

**CLAIM OF PRIORITY**

Sir:

Applicants hereby claim priority under 35 U.S.C. §119 and/or 120, from  
U.K Application No. 0101300.2 and International patent application number PCT/GB02/00215,  
a certified copy of each is enclosed.

Acknowledgment of the claim of priority and of the receipt of said certified copy  
are respectfully requested.

Respectfully submitted,

FROMMER LAWRENCE & HAUG LLP

By:

Thomas J. Kowalski  
THOMAS J. KOWALSKI, Reg. No. 32,147  
(212) 588-0800



INVESTOR IN PEOPLE

The Patent Office  
Concept House  
Cardiff Road  
Newport  
South Wales  
NP10 8QQ

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed

Dated 31 July 2005



## Page 03

P10490GB

1

## Primordial Germ Cell Genes

The present invention relates to genes which are expressed exclusively in the earliest populations of primordial germ cells (PGCs) and the use of said genes and the products thereof in identification of PGCs in cell populations.

### Introduction

Post fertilisation, the early mammalian embryo undergoes four rounds of cleavage to form a morula of 16 cells. These cells, following further rounds of division, develop into a blastocyst in which the cells can be divided into two distinct regions; the inner cell mass, which will form the embryo, and the trophectoderm, which will form extra-embryonic tissue, such as the placenta.

The cells that form part of the embryo up until the formation of the blastocyst are totipotent; in other words, each of the cells has the ability to give rise to a complete individual embryo, and to all the extra-embryonic tissues required for its development. After blastocyst formation, the cells of the inner cell mass are no longer totipotent, but are pluripotent, in that they can give rise to a range of different tissues. A known marker for such cells is the expression of the enzyme alkaline phosphatase.

Primordial germ cells (PGCs) are pluripotent cells that have the ability to differentiate into all three primary germ layers. In mammals, the PGCs migrate from the base of the allantois, through the hindgut epithelium and dorsal mesentery, to colonise the gonadal anlage. The PGC-derived cells have a characteristically low cytoplasm/nucleus ratio, usually with prominent nucleoli. PGCs may be isolated from the embryos by removing the genital ridge of the embryo, dissociating the PGCs from the gonadal anlage, and collecting the PGCs. The earliest PGC population is reported to consist of a cluster of some 40 alkaline phosphatase positive cells, found at the base of the emerging allantois. 7.25 days post-fertilisation (Ginsburg *et al.*, (1990) Development 110:521-528).

PGCs have many applications in modern biotechnology and molecular biology. They are useful in the production of transgenic animals, where embryonic germ (EG) cells derived

P10490GB

3

domain. It is a member of a multigene family, and a plurality of members of the family have been isolated from PGCs. The GCR1 polypeptide shows 89% homology to the interferon-inducible protein (sp: INIB RAT, pir:JC1241; GenPept GI:111876). It is considered that the latter polypeptide is the rat homologue of GCR1. GCR2 is a 150 amino acid polypeptide, of approximately 18kd, with a very basic pI of 9.76. The polypeptide comprises a nuclear localisation signal, and is a nuclear protein. It has no homology to any known protein.

The invention further provides polynucleotides which comprise a contiguous stretch of nucleotides from SEQ. ID. No. 1 or SEQ. ID. No. 2, or of a sequence at least 90% homologous thereto. Advantageously, this stretch of contiguous nucleotides is 50 nucleotides in length, preferably 40, 35, 30, 25, 20, 15 or 10 nucleotides in length.

The genes GCR1 and GCR2 encode novel polypeptides, the sequences of which are set forth in SEQ. ID. No. 2 and SEQ. ID. No. 4. The invention therefore encompasses polypeptides encoded by the nucleic acids according to the invention. Preferably, the polypeptides have the sequences set forth in SEQ. ID. No. 2 and SEQ. ID. No. 4.

Antibodies may be raised against the polypeptides according to the invention. In particular, antibodies may be raised against the extracellular domain of GCR1, which is a transmembrane polypeptide.

Antibodies and nucleic acids according to the invention are useful for the identification of PGCs in cell populations. The invention therefore provides a means to isolate PGCs, useful for example for the study of germ tissue development and the generation of transgenic animals, and PGCs when isolated by a method according to the invention.

Homologues of GCR1 and GCR2, such as rat interferon-inducible protein, may also be used to identify PGCs according to the present invention.

Moreover, the invention provides a method by which genes specifically expressed in PGCs may be isolated, comprising the steps of:

a) providing a population of cells containing PGCs;

P10490GB

5

Although homology can also be considered in terms of similarity (i.e. amino acid residues having similar chemical properties/functions), in the context of the present invention it is preferred to express homology in terms of sequence identity.

- 5 Homology comparisons can be conducted by eye, or more usually, with the aid of readily available sequence comparison programs. These commercially available computer programs can calculate % homology between two or more sequences.

- 10 % homology may be calculated over contiguous sequences, i.e. one sequence is aligned with the other sequence and each amino acid in one sequence directly compared with the corresponding amino acid in the other sequence, one residue at a time. This is called an "ungapped" alignment. Typically, such ungapped alignments are performed only over a relatively short number of residues (for example less than 50 contiguous amino acids).

- 15 Although this is a very simple and consistent method, it fails to take into consideration that, for example, in an otherwise identical pair of sequences, one insertion or deletion will cause the following amino acid residues to be put out of alignment, thus potentially resulting in a large reduction in % homology when a global alignment is performed. Consequently, most sequence comparison methods are designed to produce optimal alignments that take into  
20 consideration possible insertions and deletions without penalising unduly the overall homology score. This is achieved by inserting "gaps" in the sequence alignment to try to maximise local homology.

- 25 However, these more complex methods assign "gap penalties" to each gap that occurs in the alignment so that, for the same number of identical amino acids, a sequence alignment with as few gaps as possible - reflecting higher relatedness between the two compared sequences - will achieve a higher score than one with many gaps. "Affine gap costs" are typically used that charge a relatively high cost for the existence of a gap and a smaller penalty for each subsequent residue in the gap. This is the most commonly used gap scoring system. High  
30 gap penalties will of course produce optimised alignments with fewer gaps. Most alignment programs allow the gap penalties to be modified. However, it is preferred to use the default values when using such software for sequence comparisons. For example when using the

P10490GB

7

In any case, however, the key feature of the sequences of the invention – namely that they are specific for PGCs and can serve as a marker for PGCs in a cell population – is retained.

Natural variants of GCR1 and GCR2 are likely to comprise conservative amino acid substitutions. Conservative substitutions may be defined, for example according to the Table below. Amino acids in the same block in the second column and preferably in the same line in the third column may be substituted for each other:

ALIPHATIC	Non-polar	G A P
		I L V
	Polar - uncharged	C S T M
		N Q
	Polar - charged	D E
		K R
AROMATIC		H F W Y

10

Polypeptides of the invention useful as markers also include fragments of the above mentioned full length polypeptides and variants thereof, including fragments of the sequences set out in SEQ. ID. No. 2 and SEQ. ID. No. 4. Preferred fragments include those which include an epitope. Suitable fragments will be at least about 5, e.g. 10, 12, 15 or 20 amino acids in length. They may also be less than 100, 75 or 50 amino acids in length. Polypeptide fragments of the GCR proteins and allelic and species variants thereof may contain one or more (e.g. 5, 10, 15, or 20) substitutions, deletions or insertions, including conserved substitutions. Where substitutions, deletion and/or insertions occur, for example in different species, preferably less than 50%, 40% or 20% of the amino acid residues depicted in the sequence listings are altered.

20

P10490GB

9

nucleotide and -9 for each mismatch. The default gap creation penalty is -50 and the default gap extension penalty is -3 for each nucleotide.

5 The present invention also encompasses nucleotide sequences that are capable of hybridising selectively to the sequences presented herein, or any variant, fragment or derivative thereof, or to the complement of any of the above. Nucleotide sequences are preferably at least 15 nucleotides in length, more preferably at least 20, 30, 40 or 50 nucleotides in length.

10 The term "hybridisation" as used herein shall include "the process by which a strand of nucleic acid joins with a complementary strand through base pairing" as well as the process of amplification as carried out in polymerase chain reaction technologies.

15 Polynucleotides of the invention capable of selectively hybridising to the nucleotide sequences presented herein, or to their complement, will be generally at least 70%, preferably at least 80 or 90% and more preferably at least 95% or 98% homologous to the corresponding nucleotide sequences presented herein over a region of at least 20, preferably at least 25 or 30, for instance at least 40, 60 or 100 or more contiguous nucleotides.

20 The term "selectively hybridisable" means that the polynucleotide used as a probe is used under conditions where a target polynucleotide of the invention is found to hybridize to the probe at a level significantly above background. The background hybridization may occur because of other polynucleotides present, for example, in the cDNA or genomic DNA library being screening. In this event, background implies a level of signal generated by interaction between the probe and a non-specific DNA member of the library which is less  
25 than 10 fold, preferably less than 100 fold as intense as the specific interaction observed with the target DNA. The intensity of interaction may be measured, for example, by radiolabelling the probe, e.g. with <sup>32</sup>P.

30 Hybridisation conditions are based on the melting temperature (T<sub>m</sub>) of the nucleic acid binding complex, as taught in Berger and Kimmel (1987, Guide to Molecular Cloning Techniques, Methods in Enzymology, Vol 152, Academic Press, San Diego CA), and confer a defined "stringency" as explained below.

P10490GB

11

by conventional means using radioactive or non-radioactive labels, or the polynucleotides may be cloned into vectors. Such primers, probes and other fragments will be at least 15, preferably at least 20, for example at least 25, 30 or 40 nucleotides in length, and are also encompassed by the term polynucleotides of the invention as used herein. Preferred  
5 fragments are less than 500, 200, 100, 50 or 20 nucleotides in length.

Polynucleotides such as a DNA polynucleotides and probes according to the invention may be produced recombinantly, synthetically, or by any means available to those of skill in the art. They may also be cloned by standard techniques.

10

In general, primers will be produced by synthetic means, involving a step wise manufacture of the desired nucleic acid sequence one nucleotide at a time. Techniques for accomplishing this using automated techniques are readily available in the art.

15 Longer polynucleotides will generally be produced using recombinant means, for example using PCR (polymerase chain reaction) cloning techniques. This will involve making a pair of primers (e.g. of about 15 to 30 nucleotides) flanking a region of the sequence which it is desired to clone, bringing the primers into contact with mRNA or cDNA obtained from an animal or human cell, performing a polymcrasc chain reaction under conditions which bring  
20 about amplification of the desired region, isolating the amplified fragment (e.g. by purifying the reaction mixture on an agarose gel) and recovering the amplified DNA. The primers may be designed to contain suitable restriction enzyme recognition sites so that the amplified DNA can be cloned into a suitable cloning vector

## 25 D. NUCLEOTIDE VECTORS

Polynucleotides of the invention can be incorporated into a recombinant replicable vector. The vector may be used to replicate the nucleic acid in a compatible host cell. Thus in a further embodiment, the invention provides a method of making polynucleotides of the  
30 invention by introducing a polynucleotide of the invention into a replicable vector, introducing the vector into a compatible host cell, and growing the host cell under conditions which bring about replication of the vector. The vector may be recovered from the host cell. Suitable host cells include bacteria such as *E. coli*, yeast, mammalian cell

P10490GB

13

- The promoter is typically selected from promoters which are functional in mammalian cells, although prokaryotic promoters and promoters functional in other eukaryotic cells may be used. The promoter is typically derived from promoter sequences of viral or eukaryotic genes. For example, it may be a promoter derived from the genome of a cell in which expression is to occur. With respect to eukaryotic promoters, they may be promoters that function in a ubiquitous manner (such as promoters of  $\alpha$ -actin,  $\beta$ -actin, tubulin) or, alternatively, a tissue-specific manner (such as promoters of the genes for pyruvate kinase). They may also be promoters that respond to specific stimuli, for example promoters that bind steroid hormone receptors. Viral promoters may also be used, for example the Moloney murine leukaemia virus long terminal repeat (MMLV LTR) promoter, the Rous sarcoma virus (RSV) LTR promoter or the human cytomegalovirus (CMV) IE promoter.
- It may also be advantageous for the promoters to be inducible so that the levels of expression of the heterologous gene can be regulated during the life-time of the cell. Inducible means that the levels of expression obtained using the promoter can be regulated.
- In addition, any of these promoters may be modified by the addition of further regulatory sequences, for example enhancer sequences. Chimeric promoters may also be used comprising sequence elements from two or more different promoters described above.

#### E. HOST CELLS

- Vectors and polynucleotides of the invention may be introduced into host cells for the purpose of replicating the vectors/polynucleotides and/or expressing the proteins of the invention encoded by the polynucleotides of the invention. Although the proteins of the invention may be produced using prokaryotic cells as host cells, it is preferred to use eukaryotic cells, for example yeast, insect or mammalian cells, in particular mammalian cells.

Vectors/polynucleotides of the invention may introduced into suitable host cells using a

P10490GB

15

the antibody *in vivo* or *in vitro*. Such labels may be radioactive labels or radioopaque labels, such as metal particles, which are readily visualisable within an embryo or a cell mass. Moreover, they may be fluorescent labels or other labels which are visualisable on tissue samples.

5

Recombinant DNA technology may be used to improve the antibodies of the invention. Thus, chimeric antibodies may be constructed in order to decrease the immunogenicity thereof in diagnostic or therapeutic applications. Moreover, immunogenicity may be minimised by humanising the antibodies by CDR grafting [see European Patent Application 0 239 400 (Winter)] and, optionally, framework modification [EP 0 239 400].

10

Antibodies according to the invention may be obtained from animal serum, or, in the case of monoclonal antibodies or fragments thereof, produced in cell culture. Recombinant DNA technology may be used to produce the antibodies according to established procedure, in bacterial or preferably mammalian cell culture. The selected cell culture system preferably secretes the antibody product.

15

Therefore, the present invention includes a process for the production of an antibody according to the invention comprising culturing a host, e.g. *E. coli* or a mammalian cell, which has been transformed with a hybrid vector comprising an expression cassette comprising a promoter operably linked to a first DNA sequence encoding a signal peptide linked in the proper reading frame to a second DNA sequence encoding said antibody protein, and isolating said protein.

20

Multiplication of hybridoma cells or mammalian host cells *in vitro* is carried out in suitable culture media, which are the customary standard culture media, for example Dulbecco's Modified Eagle Medium (DMEM) or RPMI 1640 medium, optionally replenished by a mammalian serum, e.g. foetal calf serum, or trace elements and growth sustaining supplements, e.g. feeder cells such as normal mouse peritoneal exudate cells, spleen cells, bone marrow macrophages, 2-aminoethanol, insulin, transferrin, low density lipoprotein, oleic acid, or the like. Multiplication of host cells which are bacterial cells or yeast cells is likewise carried out in suitable culture media known in the art, for example for bacteria in medium LB, NZCYM, NZYM, NZM, Terrific Broth, SOB, SOC, 2 x YT.

25

30

P10490GB

17

For isolation of the antibodies, the immunoglobulins in the culture supernatants or in the ascitic fluid may be concentrated, e.g. by precipitation with ammonium sulphate, dialysis against hygroscopic material such as polyethylene glycol, filtration through selective membranes, or the like. If necessary and/or desired, the antibodies are purified by the customary chromatography methods, for example gel filtration, ion-exchange chromatography, chromatography over DEAE-cellulose and/or (immuno-) affinity chromatography, e.g. affinity chromatography with GCR1 or GCR2, or fragments thereof, or with Protein-A.

10 The invention further concerns hybridoma cells secreting the monoclonal antibodies of the invention. The preferred hybridoma cells of the invention are genetically stable, secrete monoclonal antibodies of the invention of the desired specificity and can be activated from deep-frozen cultures by thawing and recloning.

15 The invention also concerns a process for the preparation of a hybridoma cell line secreting monoclonal antibodies directed to GCR1 and/or GCR2, characterised in that a suitable mammal, for example a Balb/c mouse, is immunised with a one or more GCR1 or GCR2 polypeptides, or antigenic fragments thereof; antibody-producing cells of the immunised mammal are fused with cells of a suitable myeloma cell line, the hybrid cells  
20 obtained in the fusion are cloned, and cell clones secreting the desired antibodies are selected. For example spleen cells of Balb/c mice immunised with GCR1 and/or GCR2 are fused with cells of the myeloma cell line PA1 or the myeloma cell line Sp2/0-Ag14, the obtained hybrid cells are screened for secretion of the desired antibodies, and positive hybridoma cells are cloned.

25

Preferred is a process for the preparation of a hybridoma cell line, characterised in that Balb/c mice are immunised by injecting subcutaneously and/or intraperitoneally between  $10^6$  and  $10^7$  and  $10^8$  cells expressing GCR1 and/or GCR2 and a suitable adjuvant several times, e.g. four to six times, over several months, e.g. between two and four months, and  
30 spleen cells from the immunised mice are taken two to four days after the last injection and fused with cells of the myeloma cell line PA1 in the presence of a fusion promoter, preferably polyethylene glycol. Preferably the myeloma cells are fused with a three- to twentyfold excess of spleen cells from the immunised mice in a solution containing about

For the assembly of complete tetrameric immunoglobulin molecules and the expression of chimeric antibodies, the recombinant DNA inserts coding for heavy and light chain variable domains are fused with the corresponding DNAs coding for heavy and light chain constant domains, then transferred into appropriate host cells, for example after incorporation into hybrid vectors.

The invention therefore also concerns recombinant DNAs comprising an insert coding for a heavy chain murine variable domain of an antibody directed to GCR1 and/or GCR2 fused to a human constant domain  $\gamma$ , for example  $\gamma 1$ ,  $\gamma 2$ ,  $\gamma 3$  or  $\gamma 4$ , preferably  $\gamma 1$  or  $\gamma 4$ . Likewise the invention concerns recombinant DNAs comprising an insert coding for a light chain murine variable domain of an antibody directed to GCR1 and/or GCR2 fused to a human constant domain  $\kappa$  or  $\lambda$ , preferably  $\kappa$ .

In another embodiment the invention pertains to recombinant DNAs coding for a recombinant polypeptide wherein the heavy chain variable domain and the light chain variable domain are linked by way of a spacer group, optionally comprising a signal sequence facilitating the processing of the antibody in the host cell and/or a DNA coding for a peptide facilitating the purification of the antibody and/or a cleavage site and/or a peptide spacer and/or an effector molecule.

The DNA coding for an effector molecule is intended to be a DNA coding for the effector molecules useful in diagnostic or therapeutic applications. Thus, effector molecules which are toxins or enzymes, especially enzymes capable of catalysing the activation of prodrugs, are particularly indicated. The DNA encoding such an effector molecule has the sequence of a naturally occurring enzyme or toxin encoding DNA, or a mutant thereof, and can be prepared by methods well known in the art.

#### H. DETECTION OF PGCs IN CELL POPULATIONS

Polynucleotide probes or antibodies according to the invention may be used for the detection of PGCs in cell populations. As used herein, a "cell population" is any

P10490GB

21

and inserted into a vector as described above, using conventional PCR cloning and ligation techniques.

Antibodies may be labelled with any label capable of generating a signal. The signal may be any detectable signal, such as the induction of the expression of a detectable gene product. Examples of detectable gene products include bioluminescent polypeptides, such as luciferase and GFP, polypeptides detectable by specific assays, such as  $\beta$ -galactosidase and CAT, and polypeptides which modulate the growth characteristics of the host cell, such as enzymes required for metabolism such as HIS3, or antibiotic resistance genes such as G418. In a preferred aspect of the invention, the signal is detectable at the cell surface. For example, the signal may be a luminescent or fluorescent signal, which is detectable from outside the cell and allows cell sorting by FACS or other optical sorting techniques.

Preferred is the use of optical immunosensor technology, based on optical detection of fluorescently-labelled antibodies. Immunosensors are biochemical detectors comprising an antigen or antibody species coupled to a signal transducer which detects the binding of the complementary species (Rabbany *et al.*, 1994 *Crit Rev Biomed Eng* **22**:307-346; Morgan *et al.*, 1996 *Clin Chem* **42**:193-209). Examples of such complementary species include the antigen Zif 268 and the anti-Zif 268 antibody. Immunosensors produce a quantitative measure of the amount of antibody, antigen or hapten present in a complex sample such as serum or whole blood (Robinson 1991 *Biosens Bioelectron* **6**:183-191). The sensitivity of immunosensors makes them ideal for situations requiring speed and accuracy (Rabbany *et al.*, 1994 *Crit Rev Biomed Eng* **22**:307-346).

Detection techniques employed by immunosensors include electrochemical, piezoelectric or optical detection of the immunointeraction (Ghindilis *et al.*, 1998 *Biosens Bioelectron* **13**:113-131). An indirect immunosensor uses a separate labelled species that is detected after binding by, for example, fluorescence or luminescence (Morgan *et al.*, 1996 *Clin Chem* **42**:193-209). Direct immunosensors detect the binding by a change in potential difference, current, resistance, mass, heat or optical properties (Morgan *et al.*, 1996 *Clin Chem* **42**:193-209). Indirect immunosensors may encounter fewer problems due to non-specific binding (Attridge *et al.*, 1991 *Biosens Bioelectron* **6**:201-214; Morgan *et al.*, 1996 *Clin Chem* **42**:193-209).

P10490GB

23

and embryonic tissues just posterior to and above the most proximal part of the primitive streak. The cluster persists at this position at least until Early/Mid Bud stage. In the inbred 129Sv strain, the PGC cluster is found to contain a slightly larger number of the cells, which are more tightly packaged than in the C57Bl/6 strain. The 129Sv strain is used for subsequent experiments, as a better recovery of the earliest PGCs is obtained.

129Sv embryos are isolated at E7.5 in DMEM plus 10% FCS buffered with 25mM HEPES at room temperature and the developmental stage of each embryo is determined under a dissection microscope. The precise developmental stage can differ substantially even amongst embryos within the same litter. Embryos that are at the no bud or early bud (allantoic) stage are chosen for further dissection, which in part is dictated by the ease of identification of the region containing PGCs as seen under the dissection microscope. The fragment that is expected to contain the PGC cluster is cut out very precisely by means of solid glass needles. This region is dissociated into single cells using 0.25% trypsin-1mM EGTA/PBS treatment at 37°C for 10 min, followed by gentle pipetting with a mouth pipette. The dissected fragment usually contained between 250-300 cells. The procedure for cell dispersal with this gentle procedure left the visceral endoderm layer remained as an intact cellular sheet.

We picked single cells randomly from the cell suspension by a mouth pipette and put individual single cells (but avoiding generating air bubbles), into a thin-walled PCR tube containing 4µl of ice-cold cell lysis buffer (50mM Tris-HCl pH8.3, 75mM KCl, 3mM MgCl<sub>2</sub>, 0.5% NP-40, containing 80ng/ml pd(1)24, 5µg/ml prime RNase inhibitor, 324U/ml RNA guard, and 10mM each of dATP, dCTP, dGTP, and dTTP). The volume of medium carried with the single cell is less than 0.5µl. The tube is briefly centrifuged to ensure that the cell is indeed in the lysis buffer. During each separate experiment, we picked a total of 19 single cells, and left one tube without a cell, to serve as a negative control for the PCR amplification procedure. All the cells that are collected in tubes are kept on ice before starting the subsequent procedure.

30

The cells are lysed by incubating the tubes at 65°C for 1min, and then kept at room temperature for 1-2 min to allow the oligo dT to anneal to the RNA. First-strand cDNA synthesis is initiated by adding 50U of Moloney murine leukaemia virus (MMLV) and

P10490GB

25

**Example 2****Identification of PGCs by examination of the expression of marker genes**

5 The embryonic fragment which is excised theoretically contains three major components: the allantoic mesoderm, PGCs, and extraembryonic mesoderm surrounding PGCs. In order to identify the single cell cDNA of PGC origin amongst these samples, positive and negative selection of the constructed cDNAs is performed, by examining the expression of four marker genes (BMP4, TNAP, Hoxb1, and Oct4), which are known to be either  
10 expressed or repressed in various cell types in this region.

At the No/Early Bud stage, BMP4 is reported to be expressed in the emerging allantois and mesodermal components of the developing amnion, chorion, and visceral yolk sac (5). The boundary of BMP4 expression is very sharp, and the expression is completely  
15 excluded in the mesodermal region beneath the epithelial lining continuous from the amnionic mesoderm where the putative PGCs are determined. Therefore, BMP4 is used as a negative marker for the selection. Primer pairs are designed for amplifying the C terminal portion of BMP4 (5': GCC ATA CCT TGA CCC GCA GAA G, 3': AAA TGG CAC TCA GTT CAG TGG G). The PCR amplification is performed using 0.5µl of the  
20 cDNA solution as a template according to the following schedule: 95°C for 1 min, 55°C for 1 min, and 72°C for 1 min for 20 cycles. Among 83 samples tested, 57 samples show the expected size of bands, indicating expression of BMP4 these single cells. These samples are considered to be of allantoic mesodermal origin, and therefore excluded from amongst the candidates representing cells of PGC origin.

25

The expression of tissue non-specific alkaline phosphatase (TNAP), which has long been used as an early marker for PGCs (3), is then examined. Primer pairs are designed (5': CCC AAA GCA CCT TAT TTT TCT ACC, 3': TTG GCG AGT CTC TGC AAT TGG) and the same PCR reaction as above is performed. Amongst the 26 samples, 22 samples  
30 are judged to be positive for TNAP. From the alkaline phosphatase staining of the sectioned embryos, it is known that the somatic cells surrounding PGCs also express some amount of TNAP, although the level of expression is slightly lower than that in

P10490GB

27

### Example 3

#### Differential screening of single cell cDNA libraries

5 As the efficiency of the amplification of cDNA differs in each tube, it is very important to select the samples with the most efficiently amplified cDNA for the construction of libraries. The amplification of six different genes (ribosomal protein S12, intermediate filament protein vimentin,  $\beta$  tubulin-5,  $\alpha$  actin, Oct4, E-cadherin) is examined in the 10 PGC candidate samples, by Southern blot analysis. Judging from the overall profile of the amplification of all these six genes, three cDNA preparations are selected for the construction of libraries.

To obtain the maximum amount of double strand cDNA, an extension step is performed with 5 $\mu$ l of cell cDNA in 100 $\mu$ l of the PCR buffer described as above (including 1 $\mu$ l of  
15 Amplitaq) according to the following schedule: 94°C for 5min, 42°C for 5min, 72°C for 30min. The solution is extracted by phenol/chloroform treatment, and the amplified cDNAs are precipitated by ethanol, suspended in TE, and completely digested with EcoRI. The PCR primer and excess amount of dNTPs are removed by QIAGEN PCR Purification Kit, and all the purified cDNAs are run on a 2% low melting agarose gel.  
20 cDNAs above 500bp are cut and purified by QIAGEN Gel Purification Kit. The purified cDNAs are precipitated by ethanol and suspended in TE and ligated into  $\lambda$  ZAP II vector arms. The ligated vector is packaged, titered and the ratio of the successfully ligated clones is monitored by amplifying the inserts with T3 and T7 primers from 20 plaques. More than 95% of the phage are found to contain inserts.

25

The representation of the three genes, ribosomal protein S12,  $\beta$  tubulin-5, Oct4, is quantitated by screening 5000 plaques, and the library of the best quality among the three (S12 0.62%,  $\beta$  tubulin 0.4%, Oct4 0.5%) is used for the differential screening. As a comparison partner with the PGC probe, one of the most efficiently amplified  
30 surrounding somatic cell cDNA (Oct4 (+), TNAP(+/-), BMP(-), and Hoxb1(+)) is selected by the similar Southern blot analysis.

P10490GB

29

The first gene, GCR1 (Germ cell restricted-1), encodes a 137 amino acid protein. The best fit model of the EMBL program PredictProtein predicts two transmembrane domains, both N and C terminus ends being located outside. Database searches reveal a sequence match with the rat interferon-inducible protein (sp:INIB RAT, pir:JC1241) with unknown function. The GCR1 sequence appears six times in our screen, indicating high level expression in PGCs.

The second gene, GCR2, encodes a 150 amino acid protein with very basic pI (pI=9.67) and containing a nuclear localisation signal. This protein has no sequence match with any known proteins. EST database searches reveal that this sequence is only found in the preimplantation embryo and germ line (newborn ovary and female 12.5 mesonephros and gonad etc) ESTs, indicating the specificity of the expression in the totipotent or pluripotent cells.

#### Example 4

#### Identification of PGCs by screening for GCR1 and GCR2 expression

Although PGCs are identified in Example 2 by analysis of BMP4, TNAP, Hoxb1, and Oct4, no single one of these genes can be taken as a marker for the PGC state. However, both GCR1 and GCR2 may be used as such in the present invention.

The boundary of GCR2 expression in particular is well-defined, and the expression is substantially limited to PGCs. Therefore, GCR2 is used as a positive marker for the selection of PGCs. Primer pairs are designed for amplifying the C terminal portion of GCR2 (5': GCCATTCAGATGTCTCTGCAC, 3': CTCACAGCTTGAGGCTTCTAA). The PCR amplification is performed using 0.5µl of the cDNA solution obtained from PGCs in Example 1 as a template according to the following schedule: 95°C for 1 min, 55°C for 1 min, and 72°C for 1 min for 20 cycles. Among 83 samples tested, only those taken from PGCs show expression of GCR2. Hence, GCR2 is a positive marker for the PGC fate.

P10490GB

31

**References**

1. Brady, G. and Iscove, N.N. (1993). Construction of cDNA libraries from single cells. *Methods Enzymol.* 225, 611-623.

5 2. Dulac, C. and Axel, R. (1995). A novel family of genes encoding putative pheromone receptors in mammals. *Cell* 83, 195-206.

3. Ginsburg, M., Snow, M.H.L., and McLaren, A. (1990). Primordial germ cells in the mouse embryo during gastrulation. *Development* 110, 521-528.

10

4. Downs, K.M., and Davies, T. (1993). Staging of gastrulating mouse embryos by morphological landmarks in the dissecting microscope. *Development* 118, 1255-1266.

5. Lawson, K.A., Dunn, N.R., Roelen, B.A.J., Zeinstra, L.M., Davis, A.M., Wright, C.V.E., Korving, J.P.W.F.M., and Hogan, B.L.M. (1999). Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes&Dev.* 13, 424-436.

15

6. Yocm, Y.H., Fuhrmann, G., Ovitt, C.E., Brehm, A., Ohbo, K., Gross, M., Hubner, K., and Scholer, H.R. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* 1996, 881-894.

20

All publications mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described methods and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

25

30

P10490GB

33

11. A method according to claim 8, wherein the expression of the nucleic acid sequence is determined by detecting the protein product encoded thereby.

12. A method according to claim 11, wherein the protein product is detected by immunostaining.

13. An antibody specific for a polypeptide according to claim 7.

14. An antibody according to claim 13, specific for the extracellular domain of GCR1.

15. Use of an antibody according to claim 13 or claim 14 for the identification of a PGC in a population of cells.

16. A PGC when identified by a method according to any one of claims 8 to 12.

17. A method for isolating a gene specifically expressed in PGCs, comprising the steps of:

- a) providing a population of cells containing PGCs;
- b) isolating one or more PGCs therefrom and providing single-cell PGC isolates;
- c) amplifying the transcribed nucleic acid present in a single PGC;
- d) conducting a subtractive hybridisation screen to identify transcripts present in PGCs but not in somatic cells; and
- e) probing a nucleic acid library with one or more transcripts identified in d) to clone one or more genes which are specifically expressed in PGCs.

P10490GB

1

## SEQUENCE LISTING

SEQ. ID. No. 1

5 GCR1 full length

GCCGCAGAAAGGGCAGACCCGCAGCGCGCTCCATCCTTTGCCCTCCAGTGCT  
GCC'ITTGCTCCGCACCATGAACCACACTTCTCAAGCCTTCATCACCGCTGCCA  
GTGGAGGACAGCCCCCAAACACTACGAAAGAATCAAGGAAGAATATGAGGTGG  
10 CTGAGATGGGGGCACCGCACGGATCGGCTTCTGTCAGAACTACTGTGATCAA  
CATGCCCAGAGAGGTGT'CGGTGCC'GACCATGTGGTCTGGT'CCCTGTTCAATA  
CACTCTTCATGAACTTCTGCTGCCTGGGCTTCATAGCCTATGCCTACTCCGTGA  
AGTCTAGGGATCGGAAGATGGTGGGTGATGTGACTGGAGCCCAGGCCTACGC  
CTCCACTGCTAAGTGCCTGAACATCAGCACCTTGGTCC'ICAGCATCCTGATGG  
15 TTGTTATCACCA'ITGTTAGTGTCAATCATCATTGTTCTTAACGCTCAAAACCTTC  
ACACTTAATAGAGGATTCCGACTTCCGGTCCCTGAAGTGCTTCACCCCTCCGCAG  
CTGCGTCCCTCCTTGCCCCCTCCCTACACGCAGGTGTAACACTCATTTATCTATC  
CACAGTGGATTCAATAAAGTGCACTTGATAACCACC

20 SEQ. ID. No. 2

GCR2 full length

GGATCACAGACTGACTGCTAATTGGGTCTTGGTTTTAGGTCTTTTCAAAGACT  
25 AAGCAATC'ITG'ITCCGAGCTAGCTTTTGAGGCTTCTGCCCATCGCATCGCCAT  
GGAGGAACCATCAGAGAAAGT'CGACCCAATGAAGGACCCTGAAACTCCTCAG  
AAGAAAGATGAAGAGGACGCTTTGGATGATACAGACGTCCTACAACCAGAA  
ACACTAGTAAAGGTCA'IGAAAAAGCTAACCCTAAACCCCGGTGTC'AAAGCGGT  
CCGCACGCCGGCGCAGTCTACGGAACCGCAT'IGCAGCCGTACCTGTGGAGAA  
30 CAAGAGTGAAAAAATCCGGAGGGAAGTTCAAAGCGCCTTTCCCAAGAGAAG  
GGTCCGCACTTTGTTGT'CGGTGCTGAAAGACCCTATAGCAAAGATGAGAAGA  
CTTGTTCCGATTGAGCAGAGACAAAAAAGGCTCGAAGGAAATGAGTTTGAAC  
GGGACAGTGAGCCATTCAGATGTCTCT'GCACTTTCTGCCATTATCAAAGATGG